

Effective Capacity and QoS for Wireless Scheduling

Sanjay Shakkottai *Member, IEEE*

Abstract—Multi-user scheduling in a wireless context, where channel state information is exploited at the base-station can result in significant throughput gains to users. However, when Quality of Service (QoS) constraints are imposed (in the form of overflow probabilities), the benefits of multi-user scheduling are not clear. In this paper, we address this question for independent and identically distributed ON-OFF channel models, and study a “multi-user” formulation of effective capacity with QoS constraints. We consider a channel-aware greedy rule as well as the channel-aware max-queue rule, and show that these algorithms which yield the same long term throughput without QoS constraints have very different performance when QoS constraints are imposed.

Next, we study the effective capacity for varying channel burstiness. From results on multi-user scheduling, we expect the long-term throughput to grow with increasing channel burstiness. However, we show that the throughput with QoS constraints decreases with increasing channel burstiness. The intuitive justification for this is that with increasing burstiness, even though the long-term throughput increases, the channel access delay increases as well resulting in poor QoS performance.

Keywords: Large deviations, wireless channel, effective bandwidth, multiuser scheduling

I. INTRODUCTION

There has been a lot of interest over the past few years in characterizing the capacity as well as designing scheduling algorithms in a multi-user wireless context, when channel state information is available. Such a system consists of a base-station along with a collection of N mobile users. Packets arrive from the Internet to the base-station with their destinations being their corresponding mobile users. These packets are buffered at the base-station, and are delivered to the mobile users via the base-station in the following manner. The base-station delivers packets to the mobile users via a time-division multiplexing system. In such a system, time is divided into discrete slots (of the order of a milli-second [2]), and each time-slot is allocated by the base-station for data transfer to a mobile user in a dynamic manner (for example, based on user channel-access fairness considerations, QoS constraints, and/or data backlog at the base-station).

In these wireless systems, mobile users periodically send their “instantaneous” channel quality (channel state) to the base-station (via a signaling channel). The base-station is then permitted to use this information in order to dynamically allocate channel resources (time-slots) to different mobile users (for example, giving increased priority to those users whose channels are currently good). Examples of such systems include HDR over CDMA [2] and 1xEV-DO over CDMA

proposed by Qualcomm Inc., and EDGE over GSM. Much research has been focused on designing scheduling algorithms for such systems [21], [22], [1], [16], [17], [11], [12], [14], [15], [13], [8], [7].

A. Main Contributions

In this paper, we characterize the magnitude of the diversity gain when quality of service (QoS) constraints are imposed. We consider an ON-OFF channel model, which is independent and identically distributed across time and users. The main results are:

- (i) Algorithms which yield the same long term throughput without QoS constraints can have very different performance when QoS constraints (in the form of tail probabilities for queue lengths exceeding some threshold) are imposed. In this paper, we study a “multi-user” formulation for effective capacity [9] with QoS constraints. We consider a *greedy rule*, which picks the “best” channel at each time (without considering queue-lengths), as well as the *max-queue rule*, which picks the longest queue whose channel is ‘ON’. We have characterized the maximum throughput that can be supported with QoS using a large deviations framework. For the system we consider, *the long-term throughputs are the same under each of the rules*. However, we have shown that *the effective capacities (throughput with QoS constraints) of the two rules is considerably different when QoS constraints are imposed*.
- (ii) From results on multi-user scheduling (for instance, those in [23]), we would expect the long-term throughput to grow with increasing channel burstiness. However, with QoS constraints imposed, there is a significant drop in the per-user throughput, and the *throughput with QoS constraints decreases with increasing burstiness*. We have shown this by deriving an explicit (closed-form expression) upper-bound on the multi-user effective capacity for the N -user system under any scheduling algorithm. The intuitive justification for this result is that with increasing burstiness, even though there is multi-user diversity gain leading to increased channel rates per user, the channel access delay increases as well resulting in poor QoS performance. Thus, these results indicate that a system with restrictive QoS requirements cannot exploit multiuser scheduling gains.
- (iii) For the max-queue rule, using a sample-path large deviations approach [18], for the two-user case (with ON-OFF channels), we have characterized the most likely trajectories of the queue lengths at the base-station which lead to the QoS constraint being violated. This analysis connects the buffer overflow probability with a determin-

This research was supported by NSF Grants ACI-0305644, CNS-0325788 and CNS-0347400. Shorter versions of this paper appeared in the Proceedings of IEEE Symp. on Info. Theory (ISIT), 2003 and the Proceedings of the 41st Allerton Conference on Communication, Control, and Computing, Urbana, IL, October, 2003.

istic optimal control formulation, whose solution yields the buffer overflow probability in the large buffer regime. We further discuss extensions to the N -user case. These results give insight into the relative performance of the algorithms that we have considered.

From a technical viewpoint, the methods used in this paper connects queueing behavior under scheduling over a time-varying channel to a deterministic optimal control problem, whose cost function depends on the relative entropy of the channel distribution. To the best of our knowledge, such a control-theoretic view to understand wireless scheduling has not been explored in literature. Related work includes [3], [4], where the authors have studied the trade-off between power allocation and delay in a single user context. The idea of effective capacity [9], and its relation to buffer overflow in a single-user case has been used to characterize a time-varying wireless channel [24], where the authors have provided an efficient algorithm to map physical layer parameters (such as Doppler spectrum and channel distributions) to the QoS exponent and the probability that the buffer is non-empty.

II. SYSTEM MODEL

We model the wireless downlink system by a discrete-time vector channel model with N users. At each time-slot t , we denote $X_j(t)$ to be the rate at which data can be transmitted to user j if the user is chosen for transmission. For the purposes of this paper, we assume that the channel for each user is an ON-OFF process, and i.i.d. across users and in time, i.e., for each $j \in \{1, 2, \dots, N\}$ and each t ,

$$X_j(t) = \begin{cases} M & \text{w.p. } 1-p \\ 0 & \text{w.p. } p \end{cases} \quad (1)$$

Thus, in the ON state, the channel supports a rate M , and supports a rate 0 in the OFF state. Packets arrive to the base-station from the Internet and are destined to the mobile users. In this paper, we assume that the arrival process is deterministic, and at a constant rate λ that is the same to each of the users. Corresponding to each user is a queue (at the base-station) where arrivals are buffered. Let $Q_j(t)$ be queue length¹ of user j at time t . At each time t , the channel-state dependent scheduling algorithm at the base-station picks some user for transmission based on possibly the queue-lengths and the channel state information.

Fix any integer $\bar{T} > 0$. Given some $\varepsilon > 0$ and a buffer size B , the objective is to find the *maximum arrival rate per user* that the system can support such that the QoS requirement

$$\Pr(\max_{1 \leq i \leq N} Q_i(0) > B) \leq \varepsilon$$

where $Q_i(0)$ is the queue-length at time 0, and with initial conditions² $Q_i(-B\bar{T}) = 0, 1 \leq i \leq N$. A standard method for

studying this problem over wire-line networks is through the use of effective capacities [9]. Let us define $\xi_j(t)$ to equal 1 if user j is scheduled for transmission at time t and equal 0 other-wise. Thus, we have $Z_j(t) = \xi_j(t)X_j(t)$ to be the rate allocated to user j at time t .

With the above setup, for some fixed non-negative parameter θ , we have the effective capacity for each user j , $C_j(\theta)$ to be defined as

$$C_j(\theta) = \frac{1}{\theta} \lim_{n \rightarrow \infty} \frac{-1}{n} \log E \left(e^{-\theta \sum_{t=1}^n Z_j(t)} \right)$$

We comment that we restrict ourselves to scheduling policies such that the above limit exists. The above definition for effective capacity is motivated by the fact that it serves as a “surrogate” for describing the capacity of a time-varying channel when QoS requirements are imposed (see Lemma 2.1 below for a more formal description).

With this formulation, we have the following result. The proof of this Lemma is available in Appendix A.

Lemma 2.1: Let λ be the data arrival rate destined to each mobile user from the external network. Then, for \bar{T} large enough with $Q_i(-B\bar{T}) = 0, 1 \leq i \leq N$, and for each $j \in \{1, 2, \dots, N\}$, we have

$$\lambda \leq \min_{1 \leq j \leq N} C_j(\theta)$$

$$\iff \lim_{B \rightarrow \infty} \frac{-1}{B} \log \Pr(\max_{1 \leq j \leq N} Q_j(0) > B) \geq \theta$$

Thus, as a design rule, for some large buffer size B , given the QoS constraint ε and by choosing $\theta = -\log(\varepsilon)/B$, we have $\Pr(\max_j Q_j(0) > B) < \varepsilon$. Note that as $\theta \rightarrow 0$, the QoS constraint becomes loose (this corresponds to $\varepsilon \rightarrow 1$ for a fixed buffer size), and as $\theta \rightarrow \infty$, the QoS constraint becomes very strict (this corresponds to $\varepsilon \rightarrow 0$ for a fixed buffer size).

Computing the effective capacity in a multi-user scenario is very hard, as the statistics of the scheduling function $\xi_1(t)$ depends on the past scheduling decisions and also depends on various users queue lengths. For this reason, an alternate way that is sometimes employed is to use the equivalence of variational calculus and effective capacity through sample-path large deviations [18], [10]. Such techniques have been used in wireline networks, for instance to analyze Generalized Processor Sharing scheduling for a two user case [6] or a longest queue first algorithm [19].

Essentially, the variational problem asks the question: *Given some arrival rates to each of the queues and the fact that at-least one of the queues overflow at time zero, what is the most likely way that this event happens.* The variational problem solves a *deterministic* problem, where the large buffer is normalized to 1, and the initial condition of the queues is zero at time $-\infty$. Then, given a cost function which is based on the large deviation rate functions for the various processes involved [18] and constraints which are derived based on the scheduling algorithm, a deterministic optimization problem is setup to find the most likely queue length trajectories which lead to the (normalized) buffer overflow at time zero (in this case, at-least one of the queue lengths become equal to 1, the normalized buffer size at time 0, see Figure 2 for an illustration). It turns out that this problem is closely related

¹We use the following notation: At each time-slot t (i.e., time interval $[t, t+1)$, the arrivals occur at the end of the slot (i.e., $(t+1)^-$), followed by service, and then the queue length is measured. Thus, $Q_j(t)$ corresponds to the queue length after the arrivals and service have occurred in time-slot $(t-1)$.

²Note that if $\bar{T} = \infty$, then $Q_i(0)$ is the stationary queue length. In this paper, we consider the large buffer regime (i.e., $B \rightarrow \infty$), where the queue length at time $-B\bar{T}$ is initialized to zero. Thus, as $B \rightarrow \infty$, for any fixed $\bar{T} > 0$, $Q_i(0)$ converges to the stationary queue length (follows from Loynes’s formula, see also [9]). We adopt such a scaling for technical reasons in Section V-A.

to the effective capacity problem, and the solution to this problem can be used to compute the effective capacity per user. The queue length trajectories that results from the solution of the path optimization problem gives the most likely way that the buffers overflow in the actual stochastic system under the condition that the buffers are large.

In the rest of the paper, we consider two scheduling rules, namely the **greedy algorithm**, which at each time t , schedules the user with the best channel. Another rule that we consider is the **max-queue** rule, which schedules the user with the longest queue-length among the users whose channels are in the ON state. From the symmetry of the channel model and the arrival rates, it can be shown that if no QoS constraints are present, the long term throughput per user is the same under either scheduling rule. However, we will see in the following sections that if QoS constraints of the form discussed in this section are imposed, the arrival rates that can be supported dramatically change. Further, variational methods can be used to derive the most likely path of overflow.

III. AN UPPER BOUND

In this section, we derive an upper bound on the effective capacity of the channel.

Proposition 3.1: For the system under consideration, the effective capacity per user under any scheduling rule satisfies

$$\min_{1 \leq j \leq N} C_j(\theta) \leq \frac{-1}{\theta} \log \left[p^N + (1-p^N)e^{-M\theta/N} \right] \quad (2)$$

where p is defined in (1).

Proof: Recall that from Lemma 2.1, we have

$$\lambda \leq \min_{1 \leq j \leq N} C_j(\theta) \iff \lim_{B \rightarrow \infty} \frac{-1}{B} \log \Pr(\max_{1 \leq j \leq N} Q_j(0) > B) \geq \theta \quad (3)$$

Let us choose any λ satisfying (3). Now, suppose that $\sum_{i=1}^N Q_i(0) \geq NB$. Then, for some $j \in \{1, 2, \dots, N\}$, we have $Q_j(0) \geq B$. Thus, we have

$$\begin{aligned} \sum_{i=1}^N Q_i(0) \geq NB &\implies \max_{1 \leq i \leq N} Q_i(0) \geq B \text{ i.e.,} \\ \left\{ \sum_{i=1}^N Q_i(0) \geq NB \right\} &\subseteq \left\{ \max_{1 \leq i \leq N} Q_i(0) \geq B \right\} \end{aligned}$$

Thus, we have

$$\Pr\left(\sum_{i=1}^N Q_i(0) \geq NB\right) \leq \Pr\left(\max_{1 \leq i \leq N} Q_i(0) \geq B\right) \quad (4)$$

Next, let us consider an alternate work-conserving single server queue (with this queueing system denoted by SYS-2), with arrivals occurring at rate $\bar{\lambda} = N\lambda$, and queue length $\bar{Q}(t)$. The channel process for this system is defined to be an ON-OFF process $\bar{X}(t)$, defined by $\bar{X}(t) = \max_{1 \leq i \leq N} X_i(t)$. In other words,

$$\bar{X}(t) = \begin{cases} M & \text{w.p. } 1-p^N \\ 0 & \text{w.p. } p^N \end{cases}$$

Finally, let us define $\bar{B} = NB$ and $\bar{\theta} = \theta/N$. Note that this system corresponds to a queueing system where the channel resources in the multi-user system are *pooled* into a single shared resource. Now, observe that independent of any scheduling algorithm, we have $\bar{Q}(0) \leq \sum_{i=1}^N Q_i(0)$. Thus, we have

$$\begin{aligned} \bar{Q}(0) \geq \bar{B} &\implies \sum_{i=1}^N Q_i(0) \geq NB \text{ i.e.,} \\ \{\bar{Q}(0) \geq \bar{B}\} &\subseteq \left\{ \sum_{i=1}^N Q_i(0) \geq NB \right\} \end{aligned}$$

Thus, we have

$$\Pr(\bar{Q}(0) \geq \bar{B}) \leq \Pr\left(\sum_{i=1}^N Q_i(0) \geq NB\right) \quad (5)$$

From (4) and (5), it follows that

$$\Pr(\bar{Q}(0) \geq \bar{B}) \leq \Pr\left(\max_{1 \leq i \leq N} Q_i(0) \geq B\right) \quad (6)$$

Now, let $\bar{C}(\bar{\theta})$ be the effective capacity of SYS-2. It follows from [9] that we have

$$\bar{\lambda} \leq \bar{C}(\bar{\theta}) \iff \lim_{\bar{B} \rightarrow \infty} \frac{-1}{\bar{B}} \log \Pr(\bar{Q}(0) > \bar{B}) \geq \bar{\theta} \quad (7)$$

Next, we have

$$\begin{aligned} \lim_{B \rightarrow \infty} \frac{-1}{B} \log \Pr(\max_{1 \leq j \leq N} Q_j(0) > B) &\geq \theta \\ \implies \lim_{B \rightarrow \infty} \frac{-1}{B} \log \Pr(\bar{Q}(0) > NB) &\geq \theta \\ \iff \lim_{B \rightarrow \infty} \frac{-1}{NB} \log \Pr(\bar{Q}(0) > NB) &\geq \theta/N \\ \iff \lim_{\bar{B} \rightarrow \infty} \frac{-1}{\bar{B}} \log \Pr(\bar{Q}(0) > \bar{B}) &\geq \bar{\theta} \end{aligned} \quad (8)$$

From (3), (7) and (8), it now follows that

$$\lambda \leq \min_{1 \leq j \leq N} C_j(\theta) \implies \bar{\lambda} \leq \bar{C}(\bar{\theta})$$

As λ is arbitrary, it now follows from the definition of the parameters in SYS-2 that

$$\min_{1 \leq j \leq N} C_j(\theta) \leq \frac{1}{N} \bar{C}\left(\frac{\theta}{N}\right) \quad (9)$$

Finally, from standard effective capacity computations for SYS-2, the result follows. ■

From results on multi-user scheduling (for instance, those in [23]), we would expect the long term throughput to grow with increasing burstiness. However, with QoS constraints imposed, using the upper-bound, we can show that there is a significant drop in the per-user throughput, and the throughput with QoS constraints decreases with increasing burstiness (see Section VI). Thus, we conclude that for multiuser gains to be usefully exploited, the QoS requirements need to “match” the time-scale of the channel burstiness.

IV. THE GREEDY RULE

In this section, we consider the greedy scheduling rule, and derive the effective capacity under this rule.

At each time-slot t , the greedy scheduling rule is the following: Define $\mathcal{E}(t)$ to be the collection of channels which are in the ON state at time t , i.e.,

$$\mathcal{E}(t) = \arg \max_{1 \leq j \leq N} X_j(t)$$

The greedy scheduling rule at time t is to pick user $j \in \mathcal{E}$ with probability $\frac{1}{|\mathcal{E}(t)|}$, i.e., the rule is to pick equi-probably among the users whose channels are in the ON state. This rule is the greedy algorithm or the max-rate rule. In this section, we derive the effective capacity with QoS associated with this rule.

Proposition 4.1: Let the effective capacity for user j under the greedy rule be denoted by $C_j^{\text{greedy}}(\theta)$. For the system under consideration, under the greedy scheduling rule, we have the effective capacity per user to be the same for all users. For each user $j = 1, 2, \dots, N$, we have

$$C_j^{\text{greedy}}(\theta) = \frac{-1}{\theta} \log \left[1 - \frac{1-p^N}{N} + \frac{1-p^N}{N} e^{-M\theta} \right] \quad (10)$$

Proof: From the symmetry of the rule and the system model, it follows that the effective capacity for each of the N users is the same. Thus, without loss of generality, we will focus on user 1. Let us define $Z(t)$ to be the rate allocated to user 1 under the greedy rule. Then, we have

$$Z(t) = \xi_1(t) X_1(t),$$

where

$$\xi_1(t) = \begin{cases} 1 & \text{if user 1 is picked for transmission} \\ 0 & \text{otherwise} \end{cases}$$

Let us define

$$\mathcal{S}(t) = \{j \mid 2 \leq j \leq N, X_j(t) = M\}.$$

From the fact that $\{X_i(t), i = 1, 2, \dots, N\}$ are i.i.d. Bernoulli random variables, it follows that the cardinality of $\mathcal{S}(t)$ is a Binomial random variable, i.e.,

$$|\mathcal{S}(t)| \sim \text{Binomial}(N-1, 1-p)$$

Hiding the time index for notational simplicity, we now have

$$\begin{aligned} \Pr(\xi_1 = 1 \mid X_1 = M) &= \\ &= \sum_{j=0}^{N-1} \Pr(\xi_1 = 1 \mid X_1 = M, |\mathcal{S}| = j) \Pr(|\mathcal{S}| = j) \\ &= \sum_{j=0}^{N-1} \left(\frac{1}{j+1} \right) \binom{N-1}{j} (1-p)^j p^{N-1-j} \\ &= \frac{1}{N(1-p)} \sum_{j=0}^{N-1} \frac{N!}{(j+1)!(N-j-1)!} (1-p)^{j+1} p^{N-j-1} \\ &= \frac{1}{N(1-p)} \sum_{k=1}^N \binom{N}{k} (1-p)^k p^{N-k} \\ &= \frac{1-p^N}{N(1-p)} \end{aligned}$$

Thus, it follows that we have

$$\begin{aligned} \Pr(Z = 1) &= \Pr(\xi_1 X_1 = M) \\ &= \Pr(\xi_1 X_1 = M \mid X_1 = M) \Pr(X_1 = M) \\ &= \Pr(\xi_1 = 1 \mid X_1 = M) \Pr(X_1 = M) \\ &= \frac{1-p^N}{N} \end{aligned}$$

From standard computations and symmetry across users, it can now be shown that the effective capacity under the greedy rule, for each user $j = 1, 2, \dots, N$ is given by

$$C_j^{\text{greedy}}(\theta) = \frac{-1}{\theta} \log \left[1 - \frac{1-p^N}{N} + \frac{1-p^N}{N} e^{-M\theta} \right] \quad (11)$$

V. THE MAX-QUEUE RULE

We define the Max-Queue scheduling rule to be the following: At each time-slot t , the rule picks the user with the largest product of the queue-length and the channel rate, i.e.,

$$i = \arg \max_{1 \leq j \leq N} X_j(t) Q_j(t) \quad (12)$$

We note that formally speaking, we need a ‘‘tie-breaking’’ rule in case that two or more queues are exactly equal and are the largest, and all the corresponding channels are ON. We will assume that the rule then assigns the channel to the user with the smallest index.

This rule has been discussed in [21], [1], [14]. The authors have shown that this is a throughput-optimal rule, i.e., it has the largest stability region. Further, in [21] the authors show that this rule minimizes the total number of packets in the system in a stochastic ordering sense. In this paper, we are not so much interested in stability or the sum of the queue lengths but in the *per-user QoS performance* in the effective capacity sense. Given QoS requirements, we derive expressions for the throughput for each user. Unlike in the previous section where the effective capacity was directly computed, we resort to the variational approach to tackle the problem in this case. We take such an approach because the scheduling rule statistically couples the rates allocated to various users across time due to the algorithm’s (cross)-dependence on the queue lengths. In the following sections, we first derive the solution when there are only two users. We then comment on extensions to the N user case.

We first describe a optimal control problem (the variational problem) associated with the max-queue scheduling rule. Using sample-path large deviations, we then show that the buffer overflow probability for the max-queue rule can be related to the solution of the variational problem. This in turn allows us to determine the throughput the system can sustain with buffer overflow constraints.

A. Sample Path Large Deviations

In this section, we formally define the sample path large deviations problem associated with the buffer overflow problem we defined in Section II in the context of the max-queue rule.

We consider a two-user system, and denote the vector channel state (the components correspond to the channel states of the individual users) by (i, j) , which can take one of four values: $(0, 0), (0, 1), (1, 0), (1, 1)$. These correspond to the channel rate of user 1 being $M * i$ and the channel rate of user 2 being $M * j$ respectively. The index '0' corresponds to an OFF channel, and the index '1' corresponds to an ON channel.

Let us choose any finite positive integer time $T > 0$ large enough³. For each integer $B > 0$, and $t \in [-T, 0]$, define the empirical measure process (row vector with four components) $s_\mu^B(t)$ by

$$s_\mu^B(t) = \left(\frac{1}{B} \sum_{l=-BT}^{\lfloor Bt \rfloor - 1} \chi_{\{X_1(l)=0, X_2(l)=0\}} \quad \frac{1}{B} \sum_{l=-BT}^{\lfloor Bt \rfloor - 1} \chi_{\{X_1(l)=0, X_2(l)=M\}} \right. \\ \left. \frac{1}{B} \sum_{l=-BT}^{\lfloor Bt \rfloor - 1} \chi_{\{X_1(l)=M, X_2(l)=0\}} \quad \frac{1}{B} \sum_{l=-BT}^{\lfloor Bt \rfloor - 1} \chi_{\{X_1(l)=M, X_2(l)=M\}} \right) \quad (13)$$

where $\chi_{\{X_1(l)=i, X_2(l)=j\}}$ is the indicator function, i.e.,

$$\chi_{\{X_1(l)=i, X_2(l)=j\}} = \begin{cases} 1 & \text{if } X_1(l)=i \text{ and } X_2(l)=j \\ 0 & \text{otherwise} \end{cases}$$

Thus, for each $t \in [-T, 0]$, the components of $s_\mu^B(t)$ are the amount of time in the interval $[-T, t]$ that the channel is in each of the states $(0, 0), (0, 1), (1, 0)$ and $(1, 1)$ respectively.

Let us next define

$$q_i^B(t) = \frac{1}{B} Q(Bt), \quad i = 1, 2, \quad t \in [-T, 0]$$

to be the scaled queue length processes, with initial conditions $q_i^B(-T) = 0, i = 1, 2$. Observe that by this definition, we have

$$\Pr(\max_{i \in \{1, 2\}} q_i^B(0) \geq 1) = \Pr(\max_{i \in \{1, 2\}} Q_i(0) \geq B)$$

Next, let us define the channel rate process (also referred to as the empirical measure rate process) by $\bar{\phi}(t) = (\phi_{00}(t) \phi_{01}(t) \phi_{10}(t) \phi_{11}(t))$, such that $\sum_{i,j} \phi_{ij}(t) = 1$. Here, $\phi_{ij}(\cdot)$ are non-negative processes, where for any times $t_1 < t_2$,

$$\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \phi_{ij}(s) ds$$

is the fraction of time in the time interval $[t_1, t_2]$ that the channel vector is in state (i, j) . Formally, for large B ,

$$\|s_\mu^B(t_2) - s_\mu^B(t_1) - \left(\int_{t_1}^{t_2} \phi_{ij}(s) ds \right)\|_\infty \leq \frac{1}{B}$$

where $\|\cdot\|_\infty$ is the supremum norm, and

$$\left(\int_{t_1}^{t_2} \phi_{00}(s) ds \quad \int_{t_1}^{t_2} \phi_{01}(s) ds \quad \int_{t_1}^{t_2} \phi_{10}(s) ds \quad \int_{t_1}^{t_2} \phi_{11}(s) ds \right)$$

³The time is chosen such that $T > T^*$, where T^* is an optimal time solving (17), and also such that Lemma 2.1 (i.e., $T > \bar{T}$) is satisfied. In Section V-B, we will show that T^* is finite.

Given some channel rate process trajectories, we describe the queue length dynamics using a differential equation based notation. Based on the queue lengths of the two users, $q_1(t)$ and $q_2(t)$, we distinguish the dynamics of the system under the max-queue rule, henceforth referred to as MAX-QUEUE DYNAMICS, into three regions are described below:

Region A: Suppose $q_1(t) > q_2(t) > 0$. Then, we have

$$\begin{aligned} \dot{q}_1(t) &= \lambda - M(\phi_{10}(t) + \phi_{11}(t)) \\ \dot{q}_2(t) &= \lambda - M\phi_{01}(t) \end{aligned} \quad (14)$$

where M is the rate allocated in a particular state (see Section II) and λ is the arrival rate to each of the queues. Further, λ is assumed to be chosen such that the queues are stable. In other words, for a two-user system, $\lambda < \frac{M}{2}(1 - p^2)$. In Region A, as $q_1(t) > q_2(t)$, user '1' gets priority whenever the channel state is $(1, 1)$ leading to the equations above.

Region B: Suppose $q_2(t) > q_1(t) > 0$. Then, we have

$$\begin{aligned} \dot{q}_1(t) &= \lambda - M\phi_{10}(t) \\ \dot{q}_2(t) &= \lambda - M(\phi_{01}(t) + \phi_{11}(t)) \end{aligned}$$

Region C: In all cases where $q_1(t) > 0$ and $q_2(t) > 0$ (including when $q_1(t) = q_2(t) > 0$), we have

$$\dot{q}_1(t) + \dot{q}_2(t) = 2\lambda - M(\phi_{10}(t) + \phi_{01}(t) + \phi_{11}(t))$$

Similarly, we can define the dynamics over regions when one of the queue lengths is zero. For example, over a region when $q_2(t) = 0$, and $q_1(t) > 0$, the dynamics of $q_1(t)$ is given by (14). We comment that the above dynamics are analogous to that described in [6], [5]. Indeed, if we treat $\lambda - M\phi_{10}(\cdot)$ and $\lambda - M\phi_{01}(\cdot)$ as virtual "arrival processes," then, the dynamics are identical to that of the longest queue first policy with stochastic arrival processes and a stochastic channel process as described in [6], [5]. However, our interpretation of the processes in the case of time-varying channels is different, as the components of the process $\bar{\phi}(\cdot)$ are not independent (in fact, the vector process is a discrete-measure valued process). Thus, the associated variational problem will have a different cost function from that in [6], [5].

Next, let us define the (probability) vector

$$\bar{p} = (p_{00} \ p_{01} \ p_{10} \ p_{11}) \quad (15)$$

$$= (p^2 \ p(1-p) \ (1-p)p \ (1-p)^2), \quad (16)$$

where the components of the vector correspond to the probabilities that the channel is state $(0, 0), (0, 1), (1, 0)$, and $(1, 1)$ respectively. Then, for any probability vector⁴ $\bar{q} = (q_{00} \ q_{01} \ q_{10} \ q_{11})$, the relative entropy (see also [18], [10]) is defined as

$$H(\bar{q} | \bar{p}) = \sum_{i,j} q_{ij} \log \left(\frac{q_{ij}}{p_{ij}} \right)$$

With the above definitions, consider the following optimal control problem:

$$\theta^* = \inf_{T, \bar{\phi}(\cdot)} \int_{-T}^0 H(\bar{\phi}(s) | \bar{p}) ds \quad (17)$$

⁴A probability vector is such that the components are non-negative and the components sum to one.

such that $q_1(0) = 1$ or $q_2(0) = 1$, $q_1(-T) = q_2(-T) = 0$ and the queue length dynamics $q_1(t), q_2(t), t \in [-T, 0]$ are described by MAX-QUEUE DYNAMICS.

As in [6], [5], we will later show in Section V-B that this problem has a piece-wise linear solution (this property follows from the convexity of the relative entropy function $H(\cdot | \bar{p})$).

Finally, we prove the main result of this section.

Theorem 5.1:

$$\lim_{B \rightarrow \infty} \frac{-1}{B} \Pr(\max_{i \in \{1,2\}} q_i^B(0) \geq 1) = \theta^* \quad (18)$$

where θ^* solves the optimal control problem described in (17). **Proof:** This follows from a standard sample path large deviations argument [10, Chapter 5]. From the (finite-dimensional) Sanov's Theorem (Theorem 2.1.10 in [10]), as the channel process is i.i.d. (independent and identically distributed) across time, it follows that the empirical measure associated with the channel satisfies an LDP (Large Deviations Principle⁵) with rate function $H(\cdot | \bar{p})$. Now, from Mogulskii's Theorem (Theorem 5.1.2 in [10]), it follows that the empirical measure process defined in (13) satisfies an LDP with good rate function

$$I(s_\mu) = \begin{cases} \int_{-T}^0 H(\bar{\phi}(s) | \bar{p}) ds & \text{if } s_\mu \in \mathcal{AC}, s_\mu(-T) = \bar{0} \\ \infty & \text{otherwise} \end{cases}$$

where \mathcal{AC} denotes the space of absolutely continuous functions over the time interval $[-T, 0]$, $\bar{0}$ is the zero vector, and $\bar{\phi}(t) = \hat{s}_\mu(t)$ is the empirical measure rate process (the channel rate process).

Next, the queue length at time 0 is related to the empirical measure process through the dynamics described in Section V-A (see Appendix B for additional discussion). In Appendix B, we show that the mapping from the empirical measure process to the queue length process is continuous. Thus, from the Contraction Principle (Theorem 4.2.1 in [10]), we have that the rate function is given by the solution to the optimization problem (17), and the result follows. ■

The theorem relates the buffer overflow probability to the solution of the optimal control problem described in (17). Roughly, it states that the probability of buffer overflow is given by the probability of the “cheapest” channel distribution trajectory that leads to buffer overflow. The probability of this “cheapest” channel distribution trajectory is in-turn governed by the relative entropy of the empirical channel distribution with respect to the true channel distribution. Thus, the above result states that solving for the “cheapest” channel distribution trajectory (i.e., solving (17)) is sufficient to determine the (large buffer) overflow probability.

B. Mode of Overflow and the Large Deviations Exponent

In this section, we solve the optimal control problem described in (17).

⁵A Large Deviations Principle (LDP) characterizes the limiting behavior of a family of probability measures in terms of a rate function that describes the exponential rate of decay of the values that the measures assign to measurable sets, see [10], Section 1.2.

Lemma 5.1: Fix any $q \in (0.5, 1)$, and for $\bar{\phi} = (\phi_{00} \ \phi_{01} \ \phi_{10} \ \phi_{11})$, let

$$\bar{\phi}^*(1-q) = \arg \min_{\bar{\phi}} H(\bar{\phi} | \bar{p}) \quad (19)$$

such that $\sum_{i,j} \phi_{ij} = 1$, $\phi_{ij} \geq 0, i, j \in \{0, 1\}$, $\phi_{01} \geq 1-q$, and $\phi_{10} + \phi_{11} = 1-q$. Then, we have for any $q \in (0.5, \frac{1}{2-p})$,

$$\bar{\phi}^*(1-q) = (2q-1, 1-q, (1-q)p, (1-q)(1-p)) \quad (20)$$

and for any $q \in (\frac{1}{2-p}, 1)$,

$$\bar{\phi}^*(1-q) = (pq, q(1-p), (1-q)p, (1-q)(1-p)) \quad (21)$$

Proof: First, we note that the relative entropy, $H(\bar{\phi} | \bar{p})$, is convex in $\bar{\phi}$. This follows directly from the convexity of the function $y \log y$. Thus, as the cost function is convex, and the constraints form a convex set, a unique solution exists. We first observe that the constraints can be written as $\phi_{10} + \phi_{11} = 1-q$, $\phi_{00} + \phi_{01} = q$, $\phi_{01} \geq 1-q$, and $\phi_{ij} \geq 0$, and the cost function is

$$\begin{aligned} H(\bar{\phi} | \bar{p}) = & \left[\phi_{00} \log \left(\frac{\phi_{00}}{p_{00}} \right) + \phi_{01} \log \left(\frac{\phi_{01}}{p_{01}} \right) \right] + \\ & \left[\phi_{10} \log \left(\frac{\phi_{10}}{p_{10}} \right) + \phi_{11} \log \left(\frac{\phi_{11}}{p_{11}} \right) \right] \end{aligned} \quad (22)$$

From the structure of the cost function and the constraints, we observe that the minimization can be decoupled into two independent convex optimization problems, namely,

Problem Min-1:

$$\min_{\phi_{01}} \left[(q - \phi_{01}) \log \left(\frac{q - \phi_{01}}{p_{00}} \right) + \phi_{01} \log \left(\frac{\phi_{01}}{p_{01}} \right) \right] \quad (23)$$

subject to $q \geq \phi_{01} \geq 1-q$; and

Problem Min-2:

$$\min_{\phi_{10}} \left[(\phi_{10}) \log \left(\frac{\phi_{10}}{p_{10}} \right) + ((1-q) - \phi_{10}) \log \left(\frac{(1-q) - \phi_{10}}{p_{11}} \right) \right] \quad (24)$$

subject to $1-q \geq \phi_{10} \geq 0$. It can be verified that each of these two optimization problems have a convex function, and with linear constraints. Thus, we can solve using Lagrange multiplier techniques.

The problem **Min-2** can be solved by differentiation to get $\phi_{10}^* = (1-q)p$, and $\phi_{11}^* = (1-q)(1-p)$, which satisfies the non-negativity constraints.

To solve **Min-1**, we consider two cases, namely, and $\frac{1}{2-p} \leq q < 1$, and $0.5 < q < \frac{1}{2-p}$.

Case (i): When $q \geq \frac{1}{2-p}$, the constraint $\phi_{01} \geq 1-q$ is inactive in **Min-1**. To see this, we solve (24) by differentiation ignoring the constraints on ϕ_{01} and we have $\phi_{01}^* = q(1-p)$. We can check by direct computation that

$$q \geq \frac{1}{2-p} \iff q(1-p) \geq (1-q),$$

and $q(1-p) \leq q$. Thus, $\phi_{01}^* = q(1-p)$ solves **Min-1** for $q \geq \frac{1}{2-p}$.

Case (ii): We next consider the case where $0.5 < q < \frac{1}{2-p}$. We can see that in this case, the constraint $\phi_{01} \geq 1-q$ is

active (because solving **Min-1** by ignoring this constraint leads to a non-feasible solution). Thus, the solution in this case is $\phi_{01}^* = 1 - q$. Indeed, by writing the Lagrangian for **Min-1**, we can explicitly compute that $\phi_{01}^* = 1 - q$ satisfies the constraints, the complementary slackness condition, as well as leads to a positive Lagrange multiplier. Thus, we are done. ■

Lemma 5.2: Fix any $q \in (0.5, 1)$, and for $\bar{\psi} = (\psi_{00} \ \psi_{01} \ \psi_{10} \ \psi_{11})$, let

$$\bar{\psi}^*(1-q) = \arg \min_{\bar{\psi}} H(\bar{\psi} | \bar{p}) \quad (25)$$

such that $\sum_{i,j} \psi_{ij} = 1$, $\psi_{ij} \geq 0, i, j \in \{0, 1\}$, and $\psi_{01} + \psi_{10} + \psi_{11} = 2(1-q)$. Then, we have

$$\bar{\psi}^*(1-q) =$$

$$\left(1 - 2(1-q), \frac{2p(1-q)}{1+p}, \frac{2p(1-q)}{1+p}, \frac{2(1-p)(1-q)}{1+p}\right) \quad (26)$$

Proof: As in Lemma 5.1, from the convexity of $H(\cdot | \bar{p})$, and the convexity of the constraints, we can solve for the unique solution using Lagrange multipliers. First, we note that by symmetry, we have $\psi_{01}^*(1-q) = \psi_{10}^*(1-q)$. Thus, it follows that $\bar{\psi}^*(\cdot)$ is given

$$\bar{\psi}^*(1-q) = \arg \min_{\bar{\xi}} H(\bar{\xi} | \bar{p}) \quad (27)$$

where $\bar{\xi} = ((1-2(1-q)) \ a \ a \ b)$, and $2a+b=2(1-q)$. Using Lagrange multipliers, it can be shown that the optimal solution $\bar{\psi}^* = (1-2(1-q) \ a^* \ a^* \ b^*)$ satisfies

$$\frac{a^*}{p_{01}} = \frac{b^*}{p_{11}}, \quad (28)$$

$$2a^* + b^* = 2(1-q). \quad (29)$$

By solving the linear equations (28) and (29), the result follows. ■

Now, for $h \in [0, 1]$, let⁶ $\mathcal{A}(h) = \{t : 0 < \frac{\lambda-h/t}{M} < 0.5\}$. Then, for each $t \in \mathcal{A}(h)$, let us define

$$C_1(h, t) = tH\left(\bar{\phi}^*\left(\frac{\lambda-h/t}{M}\right) | \bar{p}\right) \quad (30)$$

$$C_1(h) = \inf_{t \in \mathcal{A}(h)} C_1(h, t) \quad (31)$$

where $\bar{\phi}^*(x)$ is the solution to the optimization problem in (20) and (21). Similarly, for $h \in [0, 1]$, define $\mathcal{B}(h) = \{t : 0 < \frac{\lambda-h/t}{M} < 0.5\}$. Then, for each $t \in \mathcal{B}(h)$, let us define

$$C_2(h, t) = tH\left(\bar{\psi}^*\left(\frac{\lambda-h/t}{M}\right) | \bar{p}\right) \quad (32)$$

$$C_2(h) = \inf_{t \in \mathcal{B}(h)} C_2(h, t) \quad (33)$$

where $\bar{\psi}^*(x) = \left(1 - 2x, \frac{2px}{1+p}, \frac{2px}{1+p}, \frac{2(1-p)x}{1+p}\right)$ is the solution to the optimization problem in (26). Finally, we can show that that $C_1(h, t)$ and $C_2(h, t)$ becomes unbounded as $t \rightarrow \infty$. This

implies that any optimizing t in each of $C_1(1)$ and $C_2(1)$ is finite.

Lemma 5.3: Fix any $l \in [0, 1]$ and $h \in [0, 1]$. We have

$$C_1(lh) = lC_1(h) \quad (34)$$

$$C_2(lh) = lC_2(h) \quad (35)$$

Proof: We first prove (34). By definition,

$$C_1(lh) = \inf_{t \in \mathcal{A}(lh)} tH\left(\bar{\phi}^*\left(\frac{\lambda-h/t}{M}\right) | \bar{p}\right) \quad (36)$$

First, let $t^*(h)$ be any time that solves (31), and let $t_2 = lt^*(h)$. By definition, $t^*(h)$ satisfies $0 < \frac{\lambda-h/t^*(h)}{M} < 1$. Thus, it follows that $t_2 \in \mathcal{A}(lh)$. Hence, by definition of $C_1(lh)$, we have

$$\begin{aligned} C_1(lh) &\geq t_2H\left(\bar{\phi}^*\left(\frac{\lambda-lh/t_2}{M}\right) | \bar{p}\right) \\ &= lt^*(h)H\left(\bar{\phi}^*\left(\frac{\lambda-h/t^*(h)}{M}\right) | \bar{p}\right) \\ &= lC_1(h) \end{aligned} \quad (37)$$

Next, let $t^*(lh)$ be any time that solves (36), and let $t_3 = t^*(lh)/l$. As before, it follows that $t_3 \in \mathcal{A}(h)$. Thus, by definition of $C_1(h)$, we have

$$\begin{aligned} C_1(h) &\geq t_3H\left(\bar{\phi}^*\left(\frac{\lambda-lh/t_3}{M}\right) | \bar{p}\right) \\ &= \frac{t^*(lh)}{l}H\left(\bar{\phi}^*\left(\frac{\lambda-lh/t^*(lh)}{M}\right) | \bar{p}\right) \\ &= \frac{1}{l}C_1(lh) \end{aligned} \quad (38)$$

Thus, (34) follows from (37) and (38). The proof for (35) is analogous to that of (34). ■

Theorem 5.2: Consider the optimal control problem (17). The solution θ^* is given by

$$\theta^* = \min\{C_1(1), C_2(1)\}, \quad (39)$$

where $C_1(1)$ and $C_2(1)$ are defined in (31) and (33) respectively.

Proof: The proof consists of two parts. We first show that the optimal trajectory is piece-wise linear. Using this property, we show (39).

First, suppose that the optimal trajectory $(q_1^*(t), q_2^*(t))$ satisfies the property that for some times t'_1 and t'_2 , we have $q_1^*(t'_1) = q_2^*(t'_1)$ and $q_1^*(t'_2) = q_2^*(t'_2)$. Then, we show that for all $t \in [t'_1, t'_2]$, we have $q_1^*(t) = q_2^*(t)$. We denote this as the *queue-equality property*. We prove this by contradiction.

Suppose that $q_1^*(t'_1) = q_2^*(t'_1)$ and $q_1^*(t'_2) = q_2^*(t'_2)$, are satisfied but $q_1^*(t') \neq q_2^*(t')$ for some $t' \in [t'_1, t'_2]$. Also, without loss of generality, assume that for $j = 1, 2$, $q_j^*(t'_2) > 0$ (if not, we can replace the trajectory from $-T$ to t'_2 with the zero queue trajectory and with a zero cost by using the constant mean channel rate trajectory over that interval). Choose t_1 to be the last time before t' such that $q_1^*(t_1) = q_2^*(t_1)$. Choose t_2 to be the first time after t' such that $q_1^*(t_2) = q_2^*(t_2)$ (see Figure 1). Let $\phi_{ij}(s), s \in [t_1, t_2], i, j \in \{0, 1\}$ be the corresponding channel rate trajectories. Now define $\bar{K} = (K_{00} \ K_{01} \ K_{10} \ K_{11})$, where

$$K_{ij} = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \phi_{ij}(s) ds, \quad i, j \in \{0, 1\}$$

⁶As λ is chosen such that the queueing system is stable, it can be seen that $\mathcal{A}(h)$ is not empty.

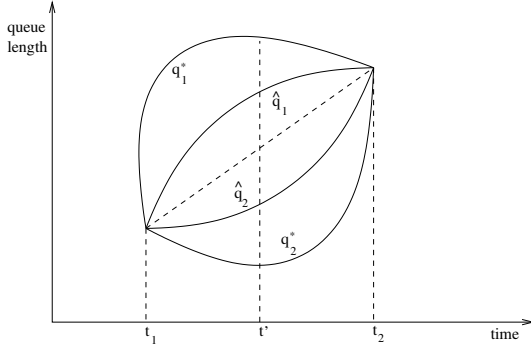


Fig. 1. The most likely mode of buffer overflow is illustrated for the max-queue rule.

and define the channel rate trajectories $\psi_{ij}(s), s \in [t_1, t_2], i, j \in \{0, 1\}$ by

$$\begin{aligned}\psi_{ij}(s) &= K_{ij} + \frac{1}{2}(\phi_{ij}(s) - K_{ij}) \\ &= \frac{1}{2}(\phi_{ij}(s) + K_{ij})\end{aligned}$$

Let $\hat{q}_1(s), \hat{q}_2(s), s \in [t_1, t_2]$ be the corresponding queue length trajectories. Thus, $\psi_{ij}(\cdot)$ are a “smoothed” version of the channel rate process, which leads to “smoothed” queue length trajectories. Note that $\psi_{ij}(\cdot)$ has been constructed such that the average time is each channel state over the interval $[t_1, t_2]$ remains unchanged. Thus, the queue lengths at time t_2 under $\psi_{ij}(\cdot)$ will remain unchanged, i.e., $\hat{q}_1(t_2) = q_1^*(t_2)$ and $\hat{q}_2(t_2) = q_2^*(t_2)$. This is illustrated in Figure 1.

We will now show that

$$\int_{t_1}^{t_2} H(\bar{\psi}(s) | \bar{p}) ds < \int_{t_1}^{t_2} H(\bar{\phi}(s) | \bar{p}) ds, \quad (40)$$

which leads to a contradiction, as $\hat{q}_1(\cdot), \hat{q}_2(\cdot), \psi_{ij}(\cdot)$ is a trajectory with lower cost, where the cost function is defined in (17). We further observe that over the interval $[t_1, t_2]$, under the contradiction hypothesis, $\phi(\cdot)$ is not constant (for if $\phi(\cdot)$ were constant, the queue lengths trajectories would be linear, and hence, equality at the end points will imply equality everywhere).

By convexity of $H(\cdot | \bar{p})$, we have

$$H(\bar{\psi}(s) | \bar{p}) \leq \frac{1}{2}H(\bar{\phi}(s) | \bar{p}) + \frac{1}{2}H(\bar{K} | \bar{p}) \quad (41)$$

Thus, we have

$$\begin{aligned}\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} H(\bar{\psi}(s) | \bar{p}) ds - \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} H(\bar{\phi}(s) | \bar{p}) ds &= \\ \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} H(\bar{\psi}(s) | \bar{p}) - H(\bar{\phi}(s) | \bar{p}) ds & \\ \leq \frac{1}{2(t_2 - t_1)} \int_{t_1}^{t_2} H(\bar{K} | \bar{p}) - H(\bar{\phi}(s) | \bar{p}) ds & \quad (42) \\ = \frac{1}{2}H(\bar{K} | \bar{p}) - \frac{1}{2(t_2 - t_1)} \int_{t_1}^{t_2} H(\bar{\phi}(s) | \bar{p}) ds, & \quad (43)\end{aligned}$$

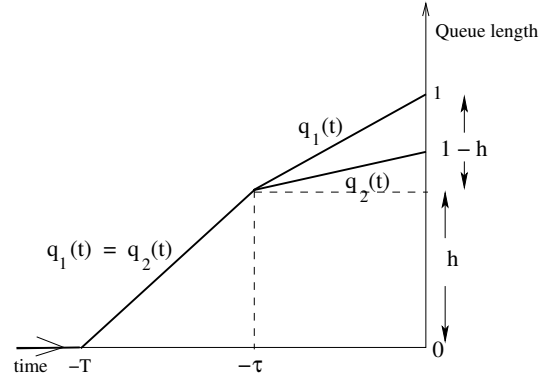


Fig. 2. Shape of the optimal queue length trajectory.

where (42) follows from (41). Applying Jensen's inequality [20], from the strict convexity of $H(\cdot | \bar{p})$, (and the fact that $\phi(\cdot)$ is not a constant trajectory) the we have

$$\begin{aligned}\frac{1}{2(t_2 - t_1)} \int_{t_1}^{t_2} H(\bar{\phi}(s) | \bar{p}) ds & \\ & > \frac{1}{2}H\left(\frac{1}{2(t_2 - t_1)} \int_{t_1}^{t_2} \bar{\phi}(s) ds | \bar{p}\right) \\ & = \frac{1}{2}H(\bar{K} | \bar{p})\end{aligned} \quad (44)$$

Thus, substituting (44) in (43), (40) follows. Thus, by contradiction, we are done.

Now, suppose that over some interval $[t_1, t_2]$, the optimal queue length trajectory $(q_1^*(t), q_2^*(t))$ satisfies $q_1^*(t) = q_2^*(t) > 0$ for all $t \in [t_1, t_2]$. Then, we next show that the optimal queue length trajectory is linear, i.e.,

$$q_i^*(t) = q_i^*(t_1) + \frac{q_i^*(t_2) - q_i^*(t_1)}{t_2 - t_1}(t - t_1) \quad i = 1, 2 \quad (45)$$

To see this, recall that for $t \in [t_1, t_2]$, from MAX-QUEUE DYNAMICS, we have

$$\dot{q}_1^*(t) + \dot{q}_2^*(t) = 2\lambda - [\phi_{11}^*(t) + \phi_{01}^*(t) + \phi_{10}^*(t)] \quad (46)$$

where $\phi_{ij}^*(\cdot), i, j \in \{0, 1\}$ is the optimal channel rate trajectory. Note that $[q_2^*(t_2) - q_2^*(t_1)] = [q_1^*(t_2) - q_1^*(t_1)]$, and let $h = [q_2^*(t_2) - q_2^*(t_1)]$. Then, from (46), we have

$$2\left(\lambda - \frac{h}{t_2 - t_1}\right) = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} [\phi_{11}^*(s) + \phi_{01}^*(s) + \phi_{10}^*(s)] ds$$

Now, from an argument identical to (44), it follows from Jensen's inequality that the optimal channel rate trajectory is constant over the interval $[t_1, t_2]$. Hence, (45) follows. Finally, an identical argument based on Jensen's inequality can be used to show that over any interval $q_1(t) \neq q_2(t)$, the optimal queue length trajectory is linear. We skip the details for brevity.

From the above discussion, it follows that the optimal queue length trajectories are piece-wise linear. Further, we now show that for some $T > 0, \tau \in [0, T]$ and $h \in [0, 1]$, the optimal queue length trajectories are of the form shown in Figure 2.

To see this, suppose $q_1^*(0) = q_2^*(0) = 1$. Also, recall that $q_1^*(-T) = q_2^*(-T) = 0$. Thus, from the queue-equality property, and piece-wise linearity of the optimal trajectories, it

follows that the optimal queue length trajectories are of the form shown in Figure 2, with $\tau = 0$ and $h = 1$.

Next, suppose the $q_1^*(0) \neq q_2^*(0)$. Without loss of generality⁷, assume that $q_1^*(0) = 1$. Let τ be the first time before 0 such that $q_1^*(-\tau) = q_2^*(-\tau)$. Thus, from the piece-wise linearity property, the queue length trajectories over the interval $[-\tau, 0]$ are linear. Further, from the queue-equality property, and piece-wise linearity of the optimal trajectories, it follows that the queue lengths are equal and the trajectory is linear over $[-T, -\tau]$.

We now derive the cost of the trajectory in Figure 2. First consider the time interval $[-\tau, 0]$, and let $\phi_{ij}(s), s \in [-\tau, 0]$ be the corresponding channel rate trajectories. As the queue length slopes are constant (with queue 1 slope being $(1 - h)/\tau$), it follows from MAX-QUEUE DYNAMICS that the channel rate trajectories are constant, and satisfy

$$\phi_{11} + \phi_{10} = \lambda - \frac{1-h}{\tau}, \quad (47)$$

$$\phi_{11} + \phi_{10} \leq \phi_{01} \quad (48)$$

Thus, the cost over this interval is given by

$$\int_{-\tau}^0 H(\bar{\phi}(s) | \bar{p}) ds = \tau H(\bar{\phi} | \bar{p})$$

where $\bar{\phi}$ satisfies (47) and (48). Thus, for a fixed h and τ , it follows from Lemma 5.1 and (30) that the minimum cost over all (constant) channel trajectories satisfying (47) and (48) is given by $C_1(1 - h, \tau)$.

Similarly, over the time interval $[-T, -\tau]$, as the queue length slopes are constant (with slope $h/(T - \tau)$), the channel rate trajectories $\psi_{ij}, i, j \in \{0, 1\}$ are constant. From MAX-QUEUE DYNAMICS, Lemma 5.2 and (32), we can show that the minimum cost and the corresponding channel rates over the interval $[-T, -\tau]$ are given by $C_2(h, T - \tau)$ and $\bar{\psi} = \left(1 - 2x, \frac{2px}{1+p}, \frac{2px}{1+p}, \frac{2(1-p)x}{1+p}\right)$ for $x = \lambda - h/(T - \tau)$ respectively.

We comment that the constraints and structure of the solution to the optimization problems Lemma 5.1 and Lemma 5.2 imply that these rates are consistent with the operation of the max-queue scheduling algorithm. In other-words, if $q_1(t) > q_2(t)$, the max-queue algorithm will allocate all channel resources to user 1 whenever its channel is ON. Thus the operation of the max-queue rule requires that $\phi_{10} + \phi_{11} \leq \phi_{01}$. This is clearly true from the solution to Lemma 5.1. A similar consistency check holds when $q_1(t) = q_2(t)$, by checking the solution from Lemma 5.2.

Thus, it follows that with the constraints in (17), from the definitions of $C_1(h)$ and $C_2(h)$, we have

$$\begin{aligned} & \inf_{T, \bar{\phi}(\cdot)} \int_{-T}^0 H(\bar{\phi}(s) | \bar{p}) ds \\ &= \inf_{\substack{h \in [0, 1] \\ T > 0, \tau \in [0, T]}} C_2((T - \tau), h) + C_1(\tau, 1 - h) \\ &= \inf_{h \in [0, 1]} C_2(h) + C_1(1 - h) \end{aligned}$$

⁷From the constraints on the optimal control problem in (17), at least one of the queue lengths equal 1 at time 0.

Next, from Lemma 5.3, we have

$$\inf_{h \in [0, 1]} [C_2(h) + C_1(1 - h)] = \inf_{h \in [0, 1]} [hC_2(1) + (1 - h)C_1(1)]$$

Thus, we have

$$\begin{aligned} \inf_{T, \bar{\phi}(\cdot)} \int_{-T}^0 H(\bar{\phi}(s) | \bar{p}) ds &= \inf_{h \in [0, 1]} [hC_1(1) + (1 - h)C_2(1)] \\ &= \min\{C_1(1), C_2(1)\} \end{aligned}$$

Thus, we are done. \blacksquare

Using Theorem 5.2 and Lemma 2.1, and given any QoS requirement θ^* , the effective capacity with the two-user greedy rule can be determined by computing the maximum arrival rate $\bar{\lambda}$ (using (39)) such that the QoS requirement of θ^* is met.

Remark 5.1: For the N user case, the formulation as a variational problem follows from a reasoning that is identical to that in the two-user case⁸. However, unlike the two-user case, we are unable to explicitly solve the resulting variational problem. We can show that the solution is piece-wise linear (the proof is similar to that of Theorem 5.2, where Jensen's inequality and convexity of the rate function is used to show this). However, we are unable to solve the resulting finite dimensional optimization in an explicit form (the analog of Lemma 5.1 and 5.2). Thus, we can reduce the infinite dimensional variational problem to a finite dimensional optimization, whose complexity grows exponentially in the number of users. We skip the details, as the results do not provide additional intuition into the system, and also leads to excessive notation.⁹

VI. NUMERICAL RESULTS AND SIMULATION

In Figure 3, we have plotted the effective capacity of the max-queue rule and the Greedy rule for various values of θ , the QoS parameter, where $\theta = -\log(\epsilon)/B$. We consider a system with two users, $p = 0.5$ and the burstiness $M = 10$. For $\theta = 0$, we observe that the effective capacities are the same for both of the rules. This is to be expected, because this corresponds to no QoS requirements being imposed. The only requirement is long term user throughput (i.e., stability). From our earlier discussion, we have seen that both of these algorithms result in the same long term throughputs (same stability conditions) if no QoS conditions are imposed. However, if QoS conditions are imposed, we can see that there is a significant difference in the per-user throughput. Thus, algorithms which yield the same long term throughput without QoS constraints (in the form of tail probabilities for queue lengths exceeding some threshold) are imposed.

In Figure 4, we consider a system with 10 users. We increase the burstiness of the channel M while keeping the mean constant by letting $1 - p = 5/M$. In other words, the mean rate in a single user system will be $M(1 - p) = 5$. We

⁸We observe that only the two users case is considered in Appendix B to prove continuity of the mapping for the contraction principle. This can be easily generalized to N users, simply by considering pair-wise differences between queue trajectories, and the rest of the proof is analogous.

⁹As a followup to this current article, recent results in [25] have in fact shown that the N user problem can be solved in closed form, and a straight-line solution holds analogous to the two-user solution derived in this paper.

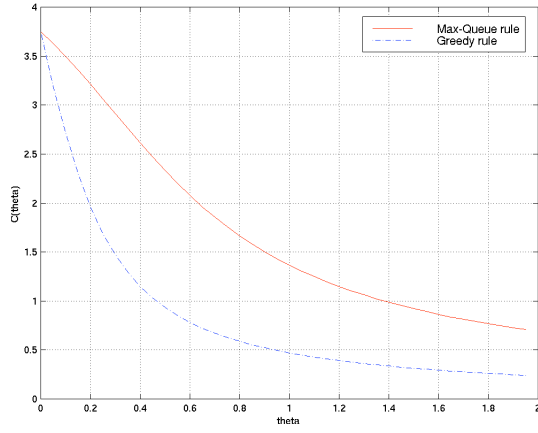


Fig. 3. Comparison of the Effective Capacity under the max-queue rule and a Greedy scheduling rule.

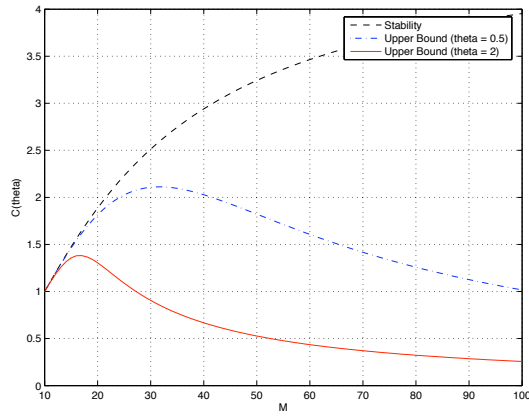


Fig. 4. Upper Bound on the Growth of Effective Capacity with channel burstiness

have plotted the user throughput when no QoS constraints are imposed (the ‘stability’ curve) as well as the upper bound on the effective capacity with $\theta = 2$ and $\theta = 0.5$. From results on multi-user scheduling (for instance, those in [23]), we would expect the long term throughput to grow with increasing burstiness. However, with QoS constraints imposed, we observe that there is a significant drop in the per-user throughput, and the throughput with QoS constraints decreases with increasing burstiness. The intuitive justification for this result is that with increasing burstiness, even though there is multi-user diversity gain leading to increased channel rates per user, the channel access delay increases as well resulting in poor QoS performance. Thus, for multiuser gains to be usefully exploited, the QoS requirements need to “match” the time-scale of the channel burstiness.

A. Simulation Results

Next, we have simulated a multi-user downlink system with bursty arrival and channel processes, which are i.i.d. across users. The arrival process is modeled by an ON-OFF Markov chain, with a packet size of 128 bytes (HDR packet size). The base-station contains 15 queues (users), each corresponding to a different data flow, and an associated scheduler. The scheduling decisions are made once every “time-slot”,

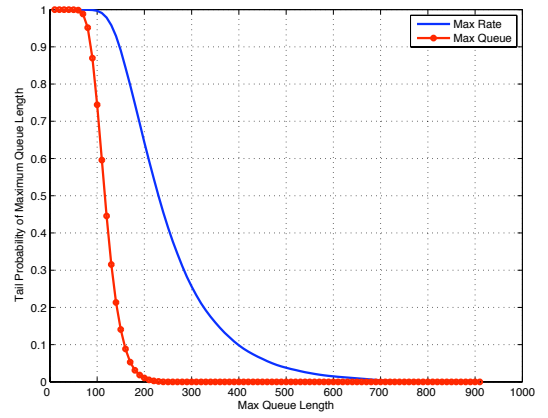


Fig. 5. Queue length distribution with the max-rate (greedy) and max-queue rules. The channel variance is 16866.

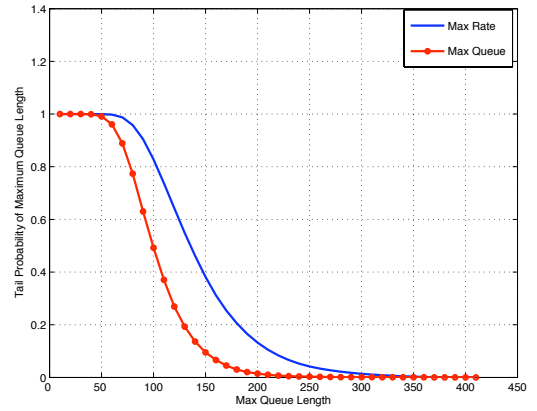


Fig. 6. Queue length distribution with the max-rate (greedy) and max-queue rules. The channel variance is 44290.7.

which is 1.667 m sec (this corresponds to the slot size in HDR). The channel process is modeled by a time-varying rate process with a support over the rate set of HDR [2] (the rates are $\{0, 8, 16, 32, 64, 256, 384, 512\}$ bytes/time-slot). By suitably choosing the distribution over this rate set of the channel, we can control the mean and variability of the channel process for each user, and hence, the burstiness of the channel (coherence time). In our simulations, the service rate has been normalized such that each users’ channel supports a mean data rate of 159 bytes/time-slot.

In Figures 5 and 6, we have plotted the delay distributions with the greedy and max-queue rules. Since the channel state is no longer binary in our simulations, we use the following generalized versions of the rules: The greedy rule is generalized to the max-rate rule, which picks the user with the highest channel rate; and the max-queue rule is generalized to picking the user who has the largest product of the channel rate and queue length [1], i.e., the largest (queue-length \times channel-rate). As we can observe from the plots, the max-queue rule outperforms the max-rate rule, which is as the analytical results suggest (in the context of ON-OFF channel models).

Next, we plot the probability that the delay of a user exceeds 500 time-slots, as a function of channel variance. In this simulation, we keep the mean channel rate fixed at

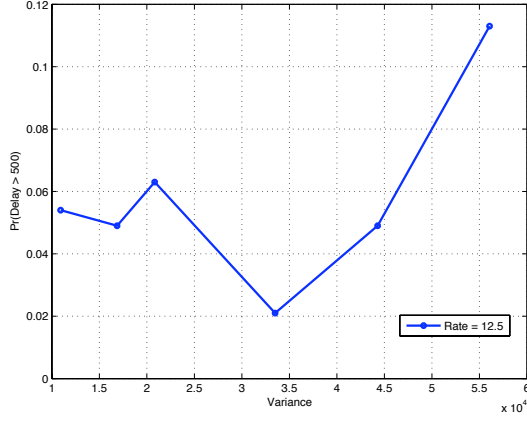


Fig. 7. Delay tail probabilities with the max-queue rules as a function of channel burstiness. The mean arrival rate of packets to each user is 12.5 bytes/time-slot.

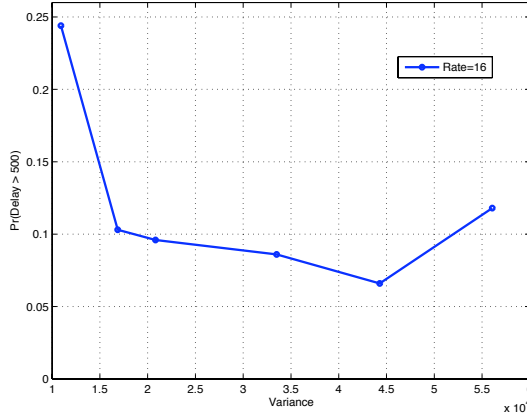


Fig. 8. Delay tail probabilities with the max-queue rules as a function of channel burstiness. The mean arrival rate of packets to each user is 16 bytes/time-slot.

159 bytes/slot. This metric is closely related to the effective capacity of the channel. In particular, the higher the channel variation, our results indicate that the effective capacity will first (possibly) increase and then decrease as the burstiness becomes large. Equivalently, the probability that the delay exceeds a threshold should first decrease and then increase with channel burstiness. This intuition is verified by the simulation results in Figures 7 and 8.

VII. CONCLUSION

Multi-user scheduling in a wireless context, where channel state information is exploited at the base-station has been shown to result in large throughput gains to users. However, when QoS constraints are imposed (in the form of overflow probabilities), the benefits of multi-user scheduling are not clear. In this paper, we have addressed this question for i.i.d. ON-OFF channel models, and study a “multi-user” formulation of effective capacity with QoS constraints. We have considered a channel-aware greedy rule as well as the channel-aware max-queue rule, and have shown that these algorithms which yield the same long term throughput without

QoS constraints have very different performance when QoS constraints are imposed.

Further, we have shown in the presence of bursty ON-OFF channels, the throughput with QoS constraints decreases with increasing channel burstiness. Thus, these results indicate that for multiuser gains to be usefully exploited, the QoS requirements need to “match” the time-scale of the channel burstiness.

ACKNOWLEDGEMENTS

The author wishes to thank Dr. Lei Ying and Prof. R. Srikant from the University of Illinois for useful comments on the proof of Theorem 5.2.

APPENDIX A: PROOF OF LEMMA 2.1

Lemma 2.1: Let λ be the data arrival rate destined to each mobile user from the external network. Then, for each $j \in \{1, 2, \dots, N\}$, we have

$$\lambda \leq \min_{1 \leq j \leq N} C_j(\theta) \iff \lim_{B \rightarrow \infty} \frac{-1}{B} \log \Pr(\max_{1 \leq j \leq N} Q_j(0) > B) \geq \theta$$

Proof: First, from the union bound, we have

$$\begin{aligned} \log \left[\Pr(\max_{1 \leq j \leq N} Q_j(0) > B) \right] &\leq \log \left[\sum_{j=1}^N \Pr(Q_j(0) > B) \right] \\ &\leq \log \left[N \max_{1 \leq j \leq N} \Pr(Q_j(0) > B) \right] \\ &= \log(N) + \max_{1 \leq j \leq N} \log [\Pr(Q_j(0) > B)], \end{aligned}$$

where the last step follows from the fact that max and log can be interchanged. Thus, we have

$$\begin{aligned} \frac{-1}{B} \log \left[\Pr(\max_{1 \leq j \leq N} Q_j(0) > B) \right] &\geq \frac{-1}{B} \log(N) + \min_{1 \leq j \leq N} \frac{-1}{B} \log [\Pr(Q_j(0) > B)] \end{aligned}$$

As N is finite, taking the limit as B goes to infinity, we have

$$\begin{aligned} \lim_{B \rightarrow \infty} \frac{-1}{B} \log \left[\Pr(\max_{1 \leq j \leq N} Q_j(0) > B) \right] &\geq \min_{1 \leq j \leq N} \lim_{B \rightarrow \infty} \frac{-1}{B} \log [\Pr(Q_j(0) > B)] \end{aligned} \quad (49)$$

Next, for any fixed i such that $1 \leq i \leq N$, we have

$$\Pr(Q_i(0) > B) \leq \Pr(\max_{1 \leq j \leq N} Q_j(0) > B)$$

Thus, we have

$$\begin{aligned} \lim_{B \rightarrow \infty} \frac{-1}{B} \log [\Pr(Q_i(0) > B)] &\geq \lim_{B \rightarrow \infty} \frac{-1}{B} \log \left[\Pr(\max_{1 \leq j \leq N} Q_j(0) > B) \right] \end{aligned}$$

As i is arbitrary, it follows that

$$\min_{1 \leq i \leq N} \lim_{B \rightarrow \infty} \frac{-1}{B} \log [\Pr(Q_i(0) > B)]$$

$$\geq \lim_{B \rightarrow \infty} \frac{-1}{B} \log \left[\Pr(\max_{1 \leq j \leq N} Q_j(0) > B) \right] \quad (50)$$

From (49) and (50), it follows that

$$\begin{aligned} & \min_{1 \leq i \leq N} \lim_{B \rightarrow \infty} \frac{-1}{B} \log [\Pr(Q_i(0) > B)] \\ &= \lim_{B \rightarrow \infty} \frac{-1}{B} \log \left[\Pr(\max_{1 \leq j \leq N} Q_j(0) > B) \right] \end{aligned} \quad (51)$$

From [9], it immediately follows that

$$\lambda \leq \min_{1 \leq j \leq N} C_j(\theta)$$

$$\iff \forall 1 \leq j \leq N, \lim_{B \rightarrow \infty} \frac{-1}{B} \log \Pr(Q_j(0) > B) \geq \theta$$

We comment that the proof in [9] considers a scenario where $Q_i(0), 1 \leq i \leq N$ are the stationary queue lengths, whereas we consider a scenario where the initial conditions are $Q_i(-B\bar{T}) = 0, 1 \leq i \leq N$. However, the proof in Theorem 3.1 in [9] immediately extends to the case with zero initial conditions at time $-B\bar{T}$, if \bar{T} is chosen large enough. By choosing $\bar{T} > 1/\alpha^*$ in Equation (7) of Theorem 3.1 in [9], the proof of Theorem 3.1 in [9] holds even for the case when $Q_i(-B\bar{T}) = 0, 1 \leq i \leq N$. Thus, we have

$$\lambda \leq \min_{1 \leq j \leq N} C_j(\theta)$$

$$\iff \min_{1 \leq j \leq N} \lim_{B \rightarrow \infty} \frac{-1}{B} \log \Pr(Q_j(0) > B) \geq \theta \quad (52)$$

Combining (52) and (51), the result follows. ■

APPENDIX B: CONTINUITY OF THE MAPPING FOR THE CONTRACTION MAPPING THEOREM

In this section, we present an argument to show that the mapping from the empirical measure process for the channel to the queue-length process, via the dynamics in Section V-A, is continuous.

To ensure continuity of the mapping, we need to show that for B large enough, a small change in the empirical measure process $s_\mu^B(t), t \in [-T, 0]$ leads to a corresponding small change in $\max\{q_1(0), q_2(0)\}$. Recall that $s_\mu^B(t), t \in [-T, 0]$ is a finite dimensional (4-dimensional in the 2 user case) increasing process. We first observe that the empirical measure process is a “stair-case” increasing process, with possible jumps of size of $1/B$, and constant over time interval of length $1/B$.

Each component of this process represents the cumulative time spent in a particular state. Thus, a small change in $s_\mu^B(t), t \in [-T, 0]$ corresponds to a small change in the *fractions* of time spent in each of the channel states.

We define the channel rate process corresponding to $s_\mu^B(t)$ by $\tilde{\phi}^B(t) = B[s_\mu^B(t) - s_\mu^B(t - 1/B)]$, where $\tilde{\phi}^B(t) = [\phi_{00}^B(t), \phi_{01}^B(t), \phi_{10}^B(t), \phi_{11}^B(t)]$. To show continuity of the mapping, we will show that to show that a small perturbation in the channel rate process, $\phi_{ij}(t), t \in [-T, 0], i, j \in \{0, 1\}$, with respect to the sup topology leads to a small change in the queue length processes, again with respect to the sup topology.

First, over any interval of time when one of the queues is zero, the system degenerates to serving the non-zero queue when ever possible (i.e., in the shared state $\phi_{11}(t)$, priority is always given to the non-empty queue). As the derivative of the queue is a linear function of $\phi_{ij}(t)$, it is easily argued that a small change in channel rate process will lead to a small change in the queue length process.

The more complex case is when both the queues are positive. To prove continuity in this case, we define $\sigma_q(t) = q_1(t) + q_2(t)$, and $\Delta_q(t) = q_1(t) - q_2(t)$. From the fact that the system is work conserving, we have¹⁰

$$\dot{\sigma}_q(t) = 2\lambda - M(1 - \phi_{00}(t))$$

Thus, it follows that σ_q is Lipschitz continuous with respect to the channel rate process. From the dynamics in Section V-A, it follows that when $\Delta_q(t) > 0$, we have $\dot{\Delta}_q(t) = \phi_{01}(t) - \phi_{10}(t) - \phi_{11}(t)$. Next, when $\Delta_q(t) < 0$, we have $\dot{\Delta}_q(t) = \phi_{01}(t) - \phi_{10}(t) + \phi_{11}(t)$.

To show that Δ_q is Lipschitz w.r.t. the empirical rate process, let us consider a perturbed channel rate process $\tilde{\phi}_{ij}$ and the corresponding queue processes $\tilde{\Delta}_q, \tilde{q}_1, \tilde{q}_2$. For each $i, j \in \{0, 1\}$, let us define $\Delta_{ij}(t) = \phi_{ij}(t) - \tilde{\phi}_{ij}(t)$.

Observe that at any time whenever Δ_q and $\tilde{\Delta}_q$ are of the opposite signs, the effect of ϕ_{11} and $\tilde{\phi}_{11}$ is to bring Δ_q and $\tilde{\Delta}_q$ toward each other. Thus, the difference between Δ_q and $\tilde{\Delta}_q$ can grow at rate at most $|\Delta_{01}(t)| + |\Delta_{10}(t)|$. On the other hand, if both Δ_q and $\tilde{\Delta}_q$ are of the same sign, the difference between Δ_q and $\tilde{\Delta}_q$ can grow at rate at most $|\Delta_{01}(t)| + |\Delta_{10}(t)| + |\Delta_{11}(t)|$. Thus, the required Lipschitz property follows.

As the sum and difference processes of the queues are Lipschitz continuous with respect to the channel rate process, it follows that the queue length processes and hence, the max of the queue length processes are Lipschitz continuous with respect to the channel rate process.

REFERENCES

- [1] M. Andrews, K. Kumaran, K. Ramanan, A. L. Stolyar, R. Vijayakumar, and P. Whiting. CDMA data QoS scheduling on the forward link with variable channel conditions, April 2000. Bell Laboratories Tech. Rep.
- [2] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, and A. Viterbi. CDMA/HDR: A bandwidth efficient high speed wireless data service for nomadic users. *IEEE Communications Magazine*, pages 70–77, July 2000.
- [3] R. Berry. Power and delay optimal transmission scheduling: Small delay asymptotics. In *IEEE International Symposium on Information Theory*, Yokohama, Japan, July 2003.
- [4] R. Berry and R. Gallager. Communication over fading channels with delay constraints. *IEEE Transactions on Information Theory*, 48(5):1135–1149, May 2002.
- [5] D. Bertsimas, I. Ch. Paschalidis, and J. N. Tsitsiklis. Asymptotic buffer overflow probabilities in multiclass multiplexers, part II: The GLQF policy, 1996. MIT LIDS Technical Report, LIDS-P-2342.
- [6] D. Bertsimas, I. Ch. Paschalidis, and J. N. Tsitsiklis. Asymptotic buffer overflow probabilities in multiclass multiplexers: An optimal control approach. *IEEE Transactions on Automatic Control*, 43:315–335, March 1998.
- [7] T. Bonald, S. C. Borst, and A. Proutiere. How mobility impacts the flow-level performance in wireless data systems. In *Proceedings of IEEE Infocom*, Hong Kong, March 2004.

¹⁰For any fixed B , we interpret the derivative as simply the scaled (by a quantity $1/B$) difference over a time-interval $[t, t + 1/B]$, i.e., $\dot{\sigma}_q(t) = B[\sigma_q(t) - \sigma_q(t - 1/B)]$.

- [8] S. C. Borst. User-level performance of channel-aware scheduling algorithms in wireless data networks. In *Proceedings of IEEE Infocom*, San Francisco, CA, June 2003.
- [9] G. de Veciana and J. Walrand. Effective bandwidths: Call admission, traffic policing and filtering for ATM networks. *Queueing Systems Theory and Applications*, 20:37–59, 1995.
- [10] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer, 2nd edition, New York, NY, 1998.
- [11] A. Eryilmaz, R. Srikant, and J.R. Perkins. Stable scheduling policies for broadcast channels. In *Proceedings of the IEEE Symposium on Information Theory*, Lausanne, Switzerland, July 2002. An earlier version appeared in *Modelling and Design of Wireless Networks, Proceedings of SPIE*, E. K. P. Chong (editor), vol. 4531, pages 70–78.
- [12] X. Liu, E. K. P. Chong, and N. B. Shroff. A framework for opportunistic scheduling in wireless networks. *To appear in the Computer Networks Journal*, 2002.
- [13] Y. Liu and E. Knightly. Opportunistic fair scheduling over multiple wireless channels. In *Proceedings of IEEE Infocom*, San Francisco, CA, June 2003.
- [14] M. J. Neely, E. Modiano, and C. E. Rohrs. Power and server allocation in a multi-beam satellite with time varying channels. In *Proceedings of IEEE Infocom*, New York, NY, June 2002.
- [15] B. Sadeghi, V. Kanodia, A. Sabharwal, and E. Knightly. Opportunistic media access for multirate ad hoc networks. In *Proceedings of ACM Mobicom 2002*, Atlanta, GA, September 2002.
- [16] S. Shakkottai and A. Stolyar. Scheduling for multiple flows sharing a time-varying channel: The exponential rule. *American Mathematical Society Translations, Series 2, A volume in memory of F. Karpelevich, Yu. M. Suhov, Editor*, 207, 2002.
- [17] S. Shakkottai and A. L. Stolyar. A study of scheduling algorithms for a mixture of real and non-real time data in HDR. In *Proceedings 17th International Teletraffic Congress*, Salvador da Bahia, Brazil, September 2001.
- [18] A. Shwartz and A. Weiss. *Large deviations for performance analysis*. Chapman and Hall, London, UK, 1995.
- [19] A. L. Stolyar and K. Ramanan. Largest weighted delay first scheduling: Large deviations and optimality. *Annals of Applied Probability*, 11(1), 2001.
- [20] D. Stroock. *Probability Theory: An Analytic View*. Cambridge University Press, 2000.
- [21] L. Tassiulas and A. Ephremides. Dynamic server allocation to parallel queues with randomly varying connectivity. *IEEE Transactions on Information Theory*, 39:466–478, March 1993.
- [22] D. Tse. Forward link multiuser diversity through proportional fair scheduling, August 1999. Presentation at Bell Labs.
- [23] P. Viswanath, D. Tse, and R. Laroia. Opportunistic beamforming using dumb antennas. *IEEE Transactions on Information Theory*, 48(6):1277–1294, June 2002.
- [24] D. Wu and R. Negi. Effective capacity: A wireless link model for support of quality of service. preprint.
- [25] Lei Ying, R. Srikant, and G. Dullerud. A large deviations analysis of scheduling in wireless networks. In *44th IEEE Conference on Decision and Control*, pages 6016 – 6021, December 2005.



Sanjay Shakkottai Sanjay Shakkottai (M'02) received his Ph.D. from the University of Illinois at Urbana-Champaign in 2002. He is with The University of Texas at Austin, where he is currently an Associate Professor in the Department of Electrical and Computer Engineering. He received the NSF CAREER award in 2004. His research interests include wireless and sensor networks, stochastic processes and queueing theory. His email address is shakkott@ece.utexas.edu.